

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 06-337695

(43)Date of publication of application : 06.12.1994

(51)Int.Cl.

G10L 3/00
G10L 3/00

(21)Application number : 05-125908

(71)Applicant : TOSHIBA CORP

(22)Date of filing : 27.05.1993

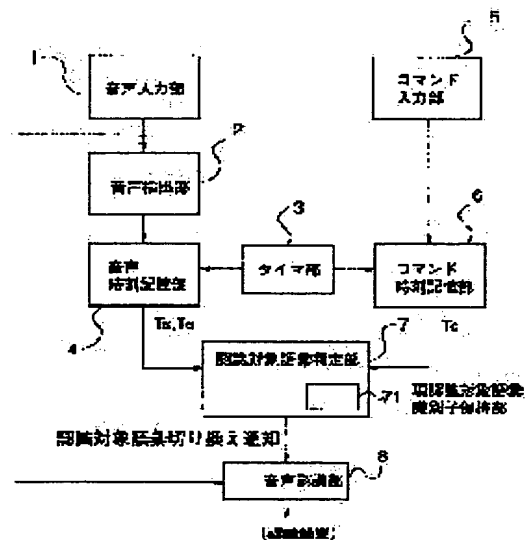
(72)Inventor : MASAI YASUYUKI

(54) VOICE RECOGNITION DEVICE

(57)Abstract:

PURPOSE: To properly switch a recognition objective vocabulary without being affected by the deviation in time due to delay, etc., at the time of transmitting a recognition objective vocabulary switching request.

CONSTITUTION: This device is constituted so that at the time of inputting the recognition objective vocabulary switching command from a command input part 5, the time when the command is inputted is read from a timer part 3 and stored in a command time storage part 6, and at the time of inputting a sound from a sound input part 1, the starting point and the ending point of the sound are detected by a sound detection part 2, and the time is read from the timer part 3 every time to be stored in a sound time storage part 4, and at the time of storing a sound input ending time, a sound input starting time stored in the sound time storage part 4 is compared with the storage contents of the command time storage part 6 by a recognition objective vocabulary decision part 7, and whether the dealing input sound is voice-recognized or not for the vocabulary after the recognition objective vocabulary is switched is decided by whether the sound input starting time is later or not, and the recognition objective vocabulary switching command is issued for a voice recognition part 8.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

BEST AVAILABLE COPY

Searching PAO
[Date of requesting appeal against examiner's
decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平6-337695

(43) 公開日 平成6年(1994)12月6日

(51) Int.Cl.⁵

G 1 0 L 3/00

識別記号

5 7 1 J
5 3 5

庁内整理番号

9379-5H
9379-5H

F I

技術表示箇所

審査請求 未請求 請求項の数 1 O L (全 8 頁)

(21) 出願番号 特願平5-125908

(22) 出願日 平成5年(1993)5月27日

(71) 出願人 000003078

株式会社東芝

神奈川県川崎市幸区堀川町72番地

(72) 発明者 正井 康之

神奈川県川崎市幸区柳町70番地 株式会社
東芝柳町工場内

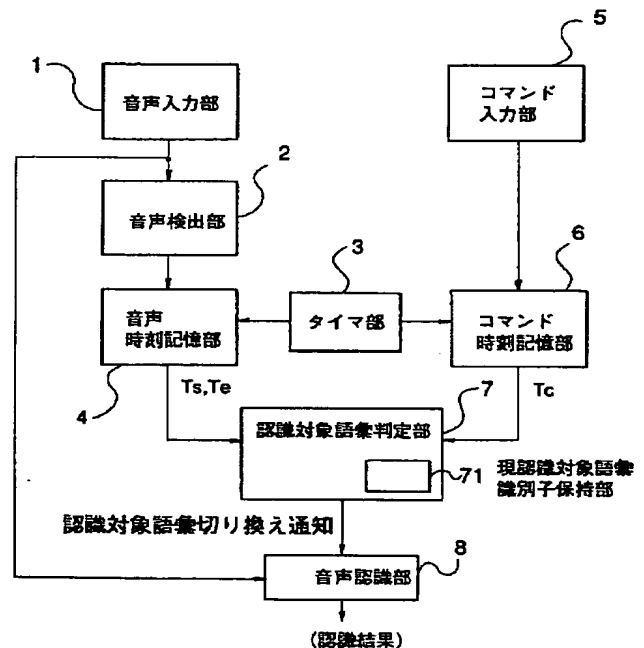
(74) 代理人 弁理士 鈴江 武彦

(54) 【発明の名称】 音声認識装置

(57) 【要約】

【目的】 認識対象語彙切り換え要求の伝達時の遅延等による時間のずれの影響を受けることなく適切に認識対象語彙を切り換えることができるようにする。

【構成】 コマンド入力部5からの認識対象語彙切り換えコマンド入力時に、その入力時点の時刻をタイマ部3から読み込みコマンド時刻記憶部6に記憶し、音声入力部1からの音声入力時には、その音声の開始点と終了点とを音声検出部2により検出して、その時刻をその都度タイマ部3から読み込んで音声時刻記憶部4に記憶し、音声入力終了時刻の記憶時に、音声時刻記憶部4に記憶済みの音声入力開始時刻とコマンド時刻記憶部6の記憶内容とを認識対象語彙判定部7にて比較して、音声入力開始時刻の方が後であるか否かにより、対応する入力音声は認識対象語彙切り換え後の語彙を対象に音声認識すべきであるか否かを判定し、音声認識部8に対して認識対象語彙の切り換えを通知する構成とする。



【特許請求の範囲】

【請求項1】入力された音声を認識する音声認識手段

と、

前記音声の入力時刻を記憶するための第1の時刻記憶手段と、

前記音声認識手段で適用する認識対象語彙の切り換え要求を入力するための認識対象語彙切り換え要求入力手段と、

前記認識対象語彙切り換え要求の入力時刻を記憶するための第2の時刻記憶手段と、

前記第1および第2の時刻記憶手段の両記憶内容と比較し、その比較結果に応じて前記音声認識手段で適用する認識対象語彙の切り換えを判定する認識対象語彙判定手段とを具備し、前記音声認識手段は、前記認識対象語彙判定手段の判定結果に従う認識対象語彙を用いて認識処理を行うことを特徴とする音声認識装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、認識対象語彙を必要に応じて切り換えて、最小限の語彙を対象に入力音声の認識を行う音声認識装置に関する。

【0002】

【従来の技術】近年、音声認識装置は、普及がめざましく、多くの分野で使用されるようになってきているものの、まだ、実用に十分な性能が得られていない。そこで、認識性能を向上させる一つの手段として、認識対象語彙を必要に応じて切り換えることで、最小限の語彙を対象に認識を行う手段が提案されている。

【0003】

【発明が解決しようとする課題】しかし、従来の認識対象語彙を切り換える手段では、スイッチ等の認識対象語彙切り換え要求入力手段から入力された認識対象語彙の切り換え要求（認識対象語彙切り換えコマンド）が当該切り換え手段（認識対象語彙判定手段）に伝達された時点を境に、音声認識装置内で認識処理を司る音声認識手段の認識対象語彙を切り換えていたため、以下に示すような問題が生じていた。

【0004】（1）認識対象語彙切り換え要求入力手段から認識対象語彙切り換え要求が入力されてから、その要求が認識対象語彙切り換え手段（認識対象語彙判定手段）に実際に伝達されるまでに、時間遅れがあると、音声認識装置の認識対象語彙が変更されていないのに、利用者が認識対象語彙変更後の語彙を発声してしまい、誤認識が生じる。

【0005】（2）音声が入力されている途中で認識対象語彙切り換え要求が入力されたとき、現在入力されている音声は、同切り換え要求の入力前の認識対象語彙で認識するのか、それとも、同切り換え要求に従って認識対象語彙切り換えを行った後の認識対象語彙で認識するのか判別がつかず、利用者の意図している認識対象語彙

での認識が行われない。

【0006】そこで本発明は、利用者が発声した音声がどの語彙を認識対象としたものであるかを判定して認識処理を行うことにより、認識対象語彙切り換え要求の伝達時の遅延等による時間のずれの影響を受けることなく適切に認識対象語彙を切り換えることができ、もって精度よく音声認識が行える音声認識装置を提供することを目的とする。

【0007】

10 【課題を解決するための手段】本発明の音声認識装置は、入力された音声を認識する音声認識手段と、上記音声の入力時刻を記憶するための第1の時刻記憶手段と、上記音声認識手段で適用する認識対象語彙の切り換え要求を入力するための認識対象語彙切り換え要求入力手段と、上記認識対象語彙切り換え要求の入力時刻を記憶するための第2の時刻記憶手段と、上記第1および第2の時刻記憶手段の両記憶内容と比較し、その比較結果に応じて音声認識手段で適用する認識対象語彙の切り換えを判定する認識対象語彙判定手段とを備え、この認識対象語彙判定手段の判定結果に従う認識対象語彙を用いて音声認識手段が認識処理を行うことを特徴とするものである。

【0008】

【作用】上記の構成においては、音声が入力された時刻が第1の時刻記憶手段に記憶され、認識対象語彙切り換え要求が入力された時刻が第2の時刻記憶手段に記憶される。これら両時刻記憶手段の記憶内容は、例えば第1の記憶手段に音声入力時刻が記憶された直後に認識対象語彙判定手段により比較される。この比較処理により、音声入力時刻には既に認識対象語彙切り換え要求が入力されているか否かが調べられ、入力された音声がどの語彙を認識対象としたものであるか否か、即ち要求された認識対象語彙の切り換えを行うか否かが判定される。音声認識手段は、この認識対象語彙判定手段の判定結果に従い、現在の認識対象語彙のまま、或いは要求された認識対象語彙に切り換えて、音声認識処理を行う。

【0009】このように、音声の入力時刻が記憶される第1の時刻記憶手段と認識対象語彙切り換え要求入力時刻が記憶される第2の時刻記憶手段の両内容を比較して、音声入力時刻には既に認識対象語彙切り換え要求が入力されているか否かを調べることにより、入力された音声がどの語彙を認識対象としたものであるか否かを判定することができるため、この判定結果に従う認識対象語彙を用いて音声認識を行うことにより、認識対象語彙切り換え要求の伝達時の遅延等による時間のずれの影響を受けることなく適切に認識対象語彙を切り換えることができ、高性能な音声認識を実現することが可能となる。

【0010】

50 【実施例】以下、本発明の一実施例について図面を参照

して説明する。図 1 は、本発明の一実施例に係る音声認識装置の構成を概略的に示すブロック図である。

【0011】この図 1 に示す音声認識装置は、マイクロホン等から入力された音声信号を増幅する音声入力部 1 と、この音声入力部 1 により増幅された音声信号から入力音声を検出する音声検出部 2 と、時計機能を持つタイマ部 3 と、音声検出部 2 により検出された入力音声の時刻をタイマ部 3 から読み込み記憶するための音声時刻記憶部 4 とを有している。

【0012】図 1 に示す音声認識装置はまた、認識対象語彙の切り換えを要求するコマンド（認識対象語彙切り換えコマンド）を入力するためのコマンド入力部 5 と、このコマンド入力部 5 により認識対象語彙切り換えコマンドが入力された時刻をタイマ部 3 から読み込み記憶するためのコマンド時刻記憶部 6 と、音声時刻記憶部 4 およびコマンド時刻記憶部 6 に記憶された各時刻を比較して入力音声の認識対象語彙切り換えコマンドの入力前か後かを判定する認識対象語彙判定部 7 と、音声認識部 8 を有している。この音声認識部 8 は、認識対象語彙判定部 7 の判定結果で決まる認識対象語彙の範囲内で、音声入力部 1 により増幅された音声信号に対する音声認識を行う。

【0013】次に、以上の構成の音声認識装置の動作を説明する。まず音声入力部 1 は、マイクロホン等から入力された音声信号を音声検出部 2 および音声認識部 8 で使用する適切な信号レベルに増幅する。

【0014】音声検出部 2 は、音声入力部 1 により増幅された音声信号を受けて、利用者が発声した音声（入力音声）を検出する。この種の音声検出の実現手段としては、従来より様々な手段が提案されている。本発明ではその方式は問わないが、入力音声信号のパワーを求め、予め設定された閾値を越えた時点を入力音声の開始点とし、閾値より小さくなった時点を入力音声の終了点として音声を検出する方式が一般的であるため、ここでは、当該方式により音声検出部 2 が実現されているものとする。なお、音声検出部 2 は、入力音声の継続時間（音声信号のパワーが閾値を連続して越えている期間）が、本装置で認識可能な語彙のうち最長の音声継続時間の語彙に対応して予め定められた時間を越えた場合にも、その時点を入力音声の終了点として検出するようになっている。

【0015】音声検出部 2 による音声検出結果は逐次音声時刻記憶部 4 に送られる。音声時刻記憶部 4 は、音声検出部 2 による音声入力開始点の検出時の時刻（音声入力開始時刻） T_s と音声入力終了点の検出時の時刻（音声入力終了時刻） T_e を、その都度タイマ部 3 から読み込み、それぞれの時刻 T_s 、 T_e を内部記憶する。なお、入力音声の時刻の記憶方法としては、上記の音声入力開始時刻と音声入力終了時刻に限ったものではなく、例えば、入力音声の中心の時刻を記憶するなど、用途に

合わせて種々変形して記憶することも可能である。

【0016】次に、図 1 の音声認識装置（内の音声認識部 8）で適用する認識対象語彙（のセット）の切り換えを要求する認識対象語彙切り換えコマンドは、コマンド入力部 5 から入力される。このコマンド入力部 5 の実現手段としては、スイッチ、キーボード等からの利用者操作によるコマンド入力、フローチャートに従った処理の流れに合わせたコマンド入力等が適用可能である。本実施例で適用される認識対象語彙切り換えコマンドは、単に認識対象語彙（のセット）の切り換えのみを指示し、切り換え後の認識対象語彙（のセット）までは指示するものでないものとする。

【0017】コマンド時刻記憶部 6 は、コマンド入力部 5 から認識対象語彙切り換えコマンドが入力されると、その入力時点の時刻（コマンド入力時刻） T_c をタイマ部 3 から読み込み、内部記憶する。

【0018】認識対象語彙判定部 7 は、このコマンド時刻記憶部 6 に記憶されたコマンド入力時刻 T_c 、および音声時刻記憶部 4 に記憶された音声入力開始時刻 T_s 、音声入力終了時刻 T_e をもとに、音声認識部 8 で適用する認識対象語彙を切り換えるか否かを判定する。この判定の詳細については後述する。なお、認識対象語彙判定部 7 は現在の認識対象語彙（のセット）の識別子を保持する現認識対象語彙識別子保持部 7 1 を有している。

【0019】音声認識部 8 は、認識対象語彙判定部 7 で判定結果に従う認識対象語彙（のセット）の範囲内で、音声入力部 1 により増幅された音声信号に基づく入力音声に対する認識処理を行う。この音声認識部 8 の実現手段としては、従来より様々な手段が提案されているが、本発明ではその方式は問わない。ここでは、例として、利用者が音声を予め登録する必要のない不特定話者音声認識方式を適用した音声認識部 8 の場合について、図 2 および図 3 を参照して説明する。

【0020】まず、音声入力部 1 により増幅された音声信号は、図 2 に示す音声認識部 8 内の音響分析部 8 1 に入力される。音響分析部 8 1 は、入力される音声信号を音響分析して特徴パラメータを求めるためのもので、ここでは、入力音声信号を LPC (Linear Predictive Coding) メルケプストラム分析する。なお、音響分析部 8 1 での音響分析は、LPC メルケプストラム分析に限るものではなく、BPF (Band Pass Filter) 分析等でもよい。

【0021】音響分析部 8 1 は、図 3 にその詳細を示すように、A/D (アナログ/デジタル) 変換器 8 1 1、パワー計算部 8 1 2 および LPC 分析部 8 1 3 から構成される。

【0022】音響分析部 8 1 に入力された音声信号は、A/D 変換器 8 1 1 にて、例えばサンプリング周波数 12 kHz、12 ビットで量子化された後、パワー計算部 8 1 2 に入力されて、その音声パワーが計算され、更に

LPC分析部813に入力されて、LPCメルケプストラム分析(LPC分析)される。このLPC分析は、例えばフレーム長16msec、フレーム周期8msecで16次のLPCメルケプストラムを分析パラメータとして行われる。

【0023】さて、図2に示す音声認識部8内には、上記の音響分析部81の他に、音響分析部81により求められた特徴パラメータによりフレーム毎にラベルを求めるための音声量子化部82、所定のPS(音声セグメント)単位の認識辞書が記憶されているPS辞書記憶部83、音声量子化部82により求められたラベル系列をHMM(Hidden Markov Model)を用いて認識するHMM認識部84、HMMのモデルMのパラメータが記憶され

$$S_{MC}^{(Ki)} = \left[\sum_{m=1}^M W_m^{(Ki)} (C \cdot \phi_m^{(Ki)})^2 \right] / \|C\|^2 \quad \dots (1)$$

なお、(1)式において、CはLPCメルケプストラム、 $W_m^{(Ki)}$ 、 $\phi_m^{(Ki)}$ はそれぞれPS名 K_i の固有値から求められる重みと固有ベクトルである。また、 (\cdot) は内積を示し、 $\| \cdot \|$ はノルムを示している。

【0026】さて、本実施例で用いられるPSとしては、例えば次のようなものがある。

- (1) 持続性セグメント : (1-1) 母音定常部
- (1-2) 摩擦子音部
- (2) 子音セグメント : 母音への渡り(過渡部)を含む部分[半音節]
- (3) 音節境界セグメント : (3-1) 母音境界
- (3-2) 母音、子音境界
- (3-3) 母音、無音境界
- (4) その他のセグメント : 無声化母音等

このうち、(1)、(2)および(4)の一部については音節を認識セグメントとする場合にも採用されることが多い。しかし、本実施例におけるPSの長所は、上記(1)、(2)、(4)に示されるセグメントに加えて上記(3)の音節境界セグメントを採用したことにある。

【0027】HMM認識部84は、音声量子化部82から出力される各フレーム毎に類似度が最大となるPS、即ちPS系列(ラベル系列)を受けて、対応する入力音声の単語照合を行う。このHMM認識部84の単語照合を以下に説明する。

【0028】本実施例における単語照合は、上記のようにPS系列をラベル系列として求め、これを単語(カテゴリ)毎のHMMに通して行うものである。ここで、HMMの一般的定式化について述べる。HMMでは、N個の状態 S_1, S_2, \dots, S_N を持ち、初期状態がこれらN個の状態に確率的に分布しているとする。音声では、一定のフレーム周期毎に、ある確率(遷移確率)で状態を遷移するモデルが使われる。遷移の際には、ある確率(出力確率)でラベルを出力するが、ラベルを出力しな

ているHMMパラメータ記憶部85、および認識結果を出力する認識結果出力部86が設けられている。

【0024】音声量子化部82は、音響分析部81で分析された特徴パラメータをPS辞書記憶部83に登録されている所定のPS単位の認識辞書と時間軸方向に連続的にマッチング処理し、各フレーム毎に類似度が最大となるPSを量子化結果として、HMM認識部84に出力する。ここで、音声量子化部82でのPSによる連続マッチング処理は、次式(1)に示す複合LPCメルケプストラム類似尺度を用いて行われる。

【0025】

【数1】

$$(Ki)^2 \bigg/ \|C\|^2 \quad \dots (1)$$

いで状態を遷移するナル遷移を導入することもある。出力ラベル系列が与えられても状態遷移系列は一意には決まらない。観測できるのは、ラベル系列だけであることからhidden(隠れ)markov model(HMM)と呼ばれている。HMMのモデルMは次の6つのパラメータから定義される。

【0029】

- N : 状態数 (状態 S_1, S_2, \dots, S_N)
- K : ラベル数 (ラベル $R=1, 2, \dots, K$)
- p_{ij} : 遷移確率 S_i から S_j に遷移する確率
- $q_{ij}(k)$: S_i から S_j への遷移の際にラベル k を出力する確率
- m_i : 初期状態確率 初期状態が S_i である確率
- F : 最終状態の集合

次に、モデルMに対して音声の特徴を反映した遷移上の制限を加える。音声では、一般的に状態 S_i から以前に通過した状態(S_{i-1}, S_{i-2}, \dots)に戻るようなループの遷移は時間的前後関係を乱すため許されない。

【0030】HMMの評価は、モデルMが第1位のラベル系列 $O_1 = o_{11}, o_{21}, \dots, o_{T1}$ を出力する確率 $P_r(O/M)$ を求めることである。認識時には、HMM認識部84で各モデルを仮定して $P_r(O/M)$ が最大になるようなモデルMをHMMパラメータ記憶部85から探す。そして、この確率 $P_r(O/M)$ が最大となるモデルに対応するカテゴリを入力音声に対する認識結果として、認識結果出力部86から出力する。

【0031】ここで、HMMパラメータ記憶部85には、認識対象語彙セットAを構成するアルファベット“a”, “b”, “c”の3語についてのモデルM(のパラメータ)と、認識対象語彙セットBを構成する数字“1”, “2”, “3”の3語についてのモデルM(のパラメータ)が記憶されている。HMM認識部84による認識時には、認識対象語彙セットAまたは認識対象語彙セットBのいずれか一方を対象に、 $P_r(O/M)$ が最

大になるようなモデルMを探す動作が行われる。いずれの認識対象語彙セットを対象とするかは、認識対象語彙判定部7により指定される。

【0032】また、HMMの学習は、予め学習データのラベル系列をHMMに与え、そこでPr(O/M)が最大となるモデルMのパラメータを推定すればよい。そして、その推定パラメータをHMMパラメータ記憶部85に登録しておく。

【0033】次に、本発明に直接関係する、認識対象語彙切り換えの動作について、認識対象語彙セットA
(“a”, “b”, “c”)と認識対象語彙セットB
(“1”, “2”, “3”)を切り換える場合を例に、具体的に説明する。

【0034】まず、現時点において、認識対象語彙判定部7の現認識対象語彙識別子保持部71には認識対象語彙セットAを示す識別子がセットされ、音声認識部8における認識対象語彙がセットAに設定されているとする。この状態で、コマンド入力部5から認識対象語彙切り換えコマンドを入力して、現在の認識対象語彙セットAから認識対象語彙セットBに切り換え設定するものとする。

【0035】ここで入力音声S1として、この入力音声S1と認識対象語彙切り換えコマンドの時間関係を図4に示す。図4において、Tsは入力音声S1の音声入力開始時刻、Teは同入力音声S1の音声入力終了時刻であり、前記したように音声時刻記憶部4に記憶されている。次にTcはコマンド(認識対象語彙切り換えコマンド)入力時刻であり、前記したようにコマンド時刻記憶部6に記憶されている。またTdはコマンド入力部5から入力された認識対象語彙切り換えコマンドが従来方式を適用したと仮定した場合に認識対象語彙判定部7に実際に伝達される時刻、Dtはその際の時間遅れである。ここで、上記の各時刻は、 $T_c < T_s < T_d < T_e$ の関係にあるものとする。

【0036】このように図4の例は、時刻Tcに認識対象語彙切り換えコマンドが入力された後に、時刻Tsから音声S1の入力が開始され、従来であれば、時刻Teに音声S1の入力が終了する前の時刻Tdに同コマンドが認識対象語彙判定部7に伝達される場合の時間関係を示している。

【0037】さて本実施例では、認識対象語彙判定部7は、音声時刻記憶部4に音声入力終了時刻Teが記憶された時点で、同記憶部4に既に記憶されている音声入力開始時刻Tsを音声の入力時刻として、コマンド時刻記憶部6に記憶されている時刻(コマンド入力時刻Tc)と比較し、音声入力開始時刻Tsの方が後であるか否かにより、対応する入力音声は認識対象語彙切り換え後の語彙を対象に音声認識すべきであるか否か、即ち認識対象語彙を切り換えるか否かを判定する。

【0038】そして認識対象語彙切り換えを判定した場

合には、認識対象語彙判定部7は、現認識対象語彙識別子保持部71の内容を、現在音声認識に適用されている認識対象語彙セットを示す識別子から、もう一方の認識対象語彙セットを示す識別子に変更し、その旨を音声認識部8(内のHMM認識部84)に通知する。

【0039】図4の例では、入力音声S1の音声入力開始時刻Tsが音声時刻記憶部4に記憶された時点には、既に認識対象語彙切り換えコマンドが入力されてそのコマンド入力時刻Tcがコマンド時刻記憶部6に記憶されている。この場合、認識対象語彙判定部7は、音声入力開始時刻Tsがコマンド入力時刻Tcより後であることから、認識対象語彙切り換えを判定し、現認識対象語彙識別子保持部71の内容を、現在の認識対象語彙セットAを示す識別子からもう一方の認識対象語彙セットBを示す識別子に変更し、音声認識部8(内のHMM認識部84)に対して認識対象語彙セットBへの切り換えを通知する。そして認識対象語彙判定部7は、この切り換え通知を行うと、コマンド時刻記憶部6の内容を無効状態に設定する。なお、音声入力開始時刻Tsがコマンド入力時刻Tcより前の場合、あるいはコマンド時刻記憶部6の内容が無効状態の場合には、音声認識部8(内のHMM認識部84)への切り換え通知はない。

【0040】音声認識部8内のHMM認識部84は、入力音声S1の入力終了時点(所定時間を経過しても音声入力が続いている場合には、その所定時間を越えた時点)から、その音声S1の認識処理を開始する。ここで、認識対象語彙判定部7から上記の切り換え通知を受けた場合には、入力音声S1の認識処理、即ちPr(O/M)が最大になるようなモデルMを探す処理を、図4に示すように、現在の認識対象語彙セットAではなくて、認識対象語彙セットBを対象に行う。これにより現在の認識対象語彙セットAを認識対象語彙セットBに切り換えるために認識対象語彙切り換えコマンドを入力してから、例えば数字“1”を発声した場合には、この数字“1”の入力音声(S1)は、認識対象語彙セットB(“1”, “2”, “3”)を対象に認識処理されるので、正しく認識される。この認識結果は、音声認識部8の認識結果出力部86から出力される。

【0041】これに対し、従来のように、認識対象語彙切り換えコマンドが認識対象語彙判定部7に伝達された時刻Tdを境に認識対象語彙を切り換える方式では、入力音声S1は、図4に示すように認識対象語彙セットA(“a”, “b”, “c”)を対象に認識処理される。このため、上記の例のように、現在の認識対象語彙セットAを認識対象語彙セットBに切り換えるために認識対象語彙切り換えコマンドを入力し、しかる後に数字“1”を発声したとしても、この数字“1”の入力音声(S1)が認識対象語彙セットA(“a”, “b”, “c”)を対象に認識処理される従来方式では、認識結果としては、アルファベット“a”, “b”, “c”の

いずれかししか得られず、正しく認識されることはない。

【0042】以上に実施例を挙げて本発明を説明したが、本発明は前記実施例に限るものではない。即ち、本発明の要旨とするところは、認識対象語彙を適切に切り換えて高い認識性能を得るために、音声の入力時刻と認識対象語彙切り換えコマンド（認識対象語彙切り換え要求）の入力時刻とを記憶して、これらの時刻から入力された音声がいずれの語彙を対象に認識処理すべきかを判定し、音声認識を行うことにある。したがって、音声認識方式、音声検出方式、認識対象語彙切り換えコマンド入力方式、認識対象語彙判定規則などは、前記実施例で示したものに限らない。

【0043】例えば、前記実施例では、認識対象語彙判定部7から音声認識部8に対する認識対象語彙の切り換え通知を、同認識部8内のHMM認識部84に導き、通知された認識対象語彙の範囲内で認識処理を行うものとして説明したが、当該切り換え通知を認識結果出力部86に導くようにし、HMM認識部84では、認識対象語彙セットAおよび認識対象語彙セットBの両セットを全て対象にして認識処理を行い、認識結果出力部86において、切り換え通知により指定された切り換え後の認識対象語彙セットの範囲内で認識結果を出力するようにすることも可能である。したがって、特許請求の範囲に記載の「認識対象語彙を用いて認識処理を行う」ことには、この認識結果の出力処理も含まれる。

【0044】また、前記実施例では、音声入力開始時刻 T_s を音声の入力時刻として、コマンド時刻記憶部6の内容（コマンド入力時刻 T_c ）と比較する場合について説明したが、音声入力開始時刻 T_s に代えて、音声入力開始時刻 T_s と音声入力終了時刻 T_e の中心の時刻（ $T_s + T_e$ ）/2、または音声入力終了時刻 T_e を音声の入力時刻とするようにしても構わない。

【0045】また、前記実施例では、音声入力終了時刻 T_e が音声時刻記憶部4に記憶された際に、この音声時刻記憶部4に既に記憶されている音声入力開始時刻 T_s を音声の入力時刻としてコマンド時刻記憶部6の内容（コマンド入力時刻 T_c ）と比較することにより認識対象語彙の切り換えを判定するものとして説明したが、こ

の例のように音声の入力時刻として音声入力開始時刻 T_s を適用する場合には、当該音声入力開始時刻 T_s が音声時刻記憶部4に記憶された際に、認識対象語彙の切り換えを判定するようにしても構わない。

【0046】また、前記実施例では、コマンド入力部5からの認識対象語彙切り換えコマンドが、認識対象語彙（のセット）の単なる切り換えのみを指示するものであるとしたが、切り換え後の認識対象語彙（セット）を併せて指定するものであっても構わない。この場合には、認識対象語彙を一定順序で切り換えるのではなく、指定の認識対象語彙に切り換えればよい。その他、本発明はその要旨を逸脱しない範囲で種々変形して実施することができる。

【0047】

【発明の効果】以上説明したように本発明によれば、音声の入力時刻および認識対象語彙の切り換え要求の入力時刻を記憶しておき、その記憶しておいた時刻から、入力された音声がいずれの語彙を認識対象語彙として発声されたものかを判定する構成としたので、認識対象語彙の切り換え要求の伝達時の遅延時間等による時間のずれの影響を受けることなく、認識対象語彙を適切に切り換えることができ、高性能な音声認識を実現することができる等の実用上多大なる効果が奏せられる。

【図面の簡単な説明】

【図1】本発明の一実施例に係る音声認識装置の基本構成を示すブロック図。

【図2】図1中の音声認識部8の構成を示すブロック図。

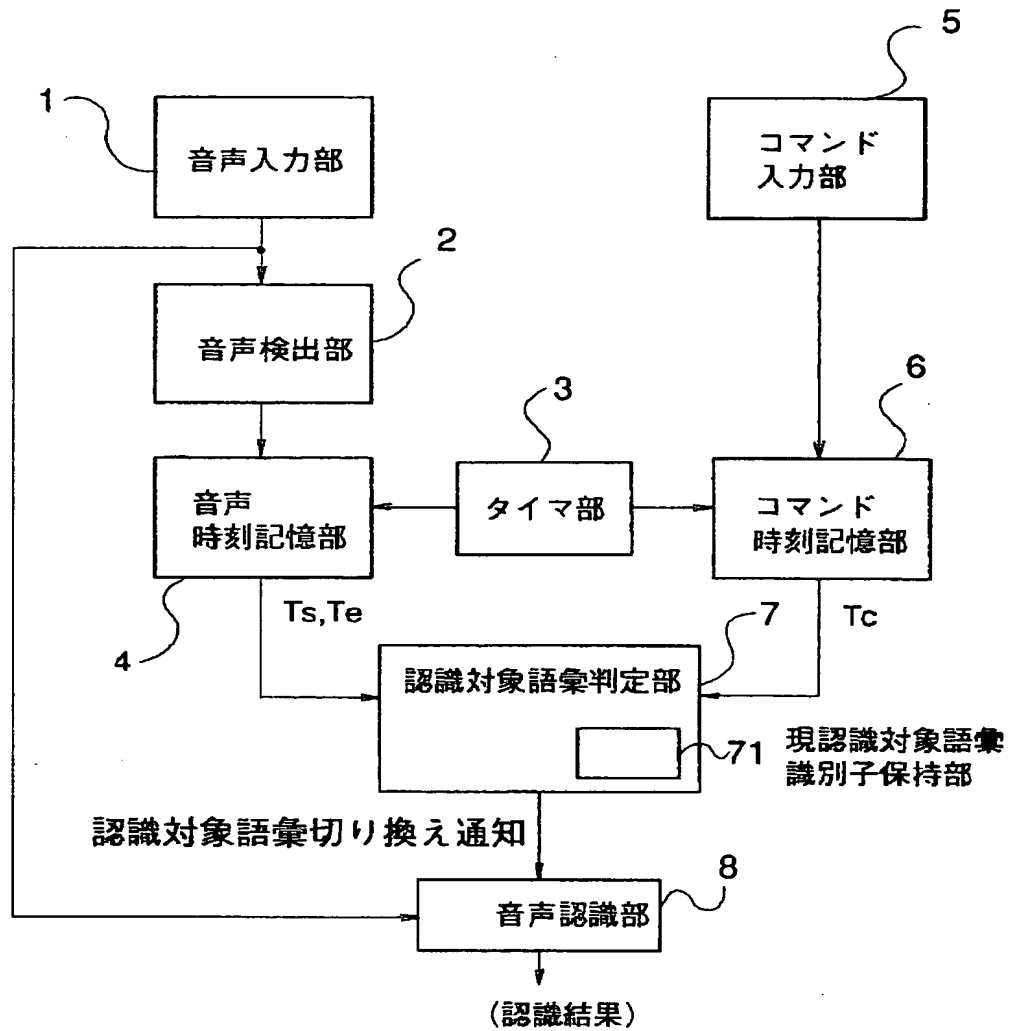
【図3】図2中の音響分析部81の詳細を示す図。

【図4】同実施例における認識対象語彙判定に使用する時刻の前後関係の一例を示す図。

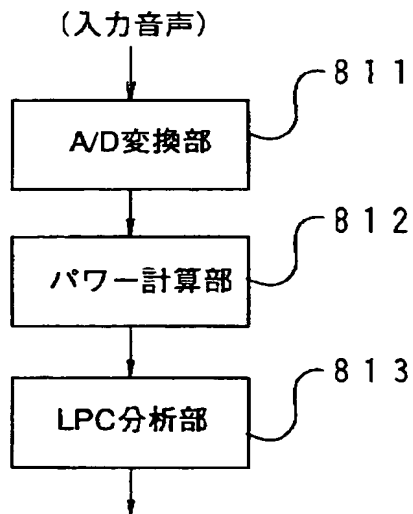
【符号の説明】

1…音声入力部、2…音声検出部、3…タイマ部、4…音声時刻記憶部（第1の時刻記憶手段）、5…コマンド入力部（認識対象語彙切り換え要求入力手段）、6…コマンド時刻記憶部（第2の時刻記憶手段）、7…認識対象語彙判定部、8…音声認識部、71…現認識対象語彙識別子保持部、A、B…認識対象語彙セット。

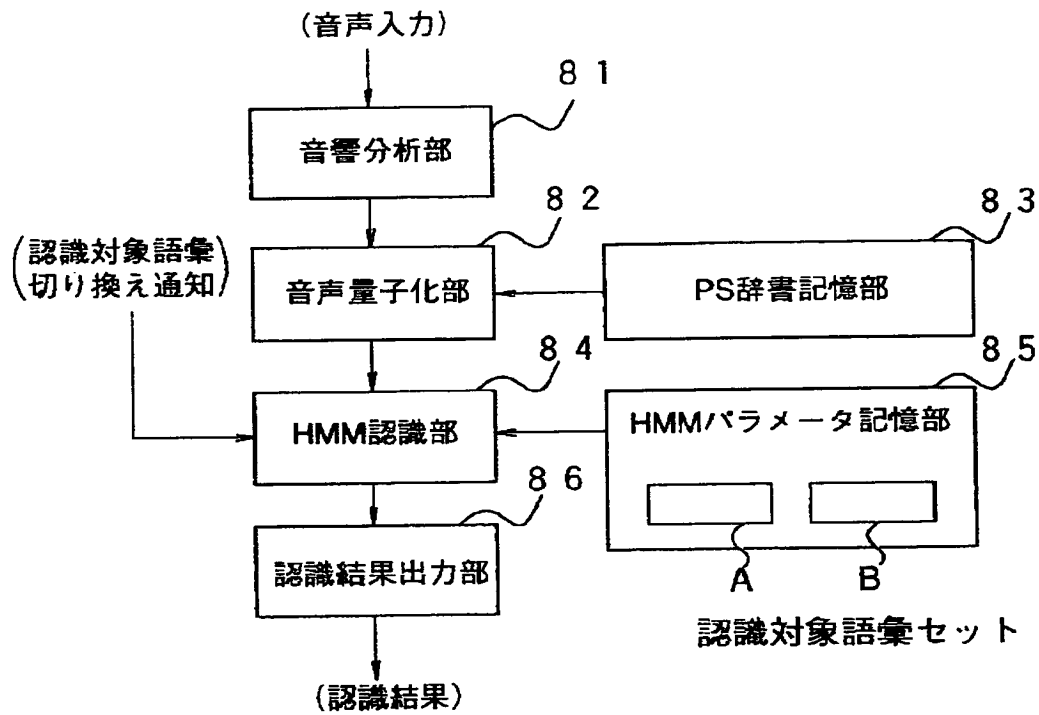
【図1】



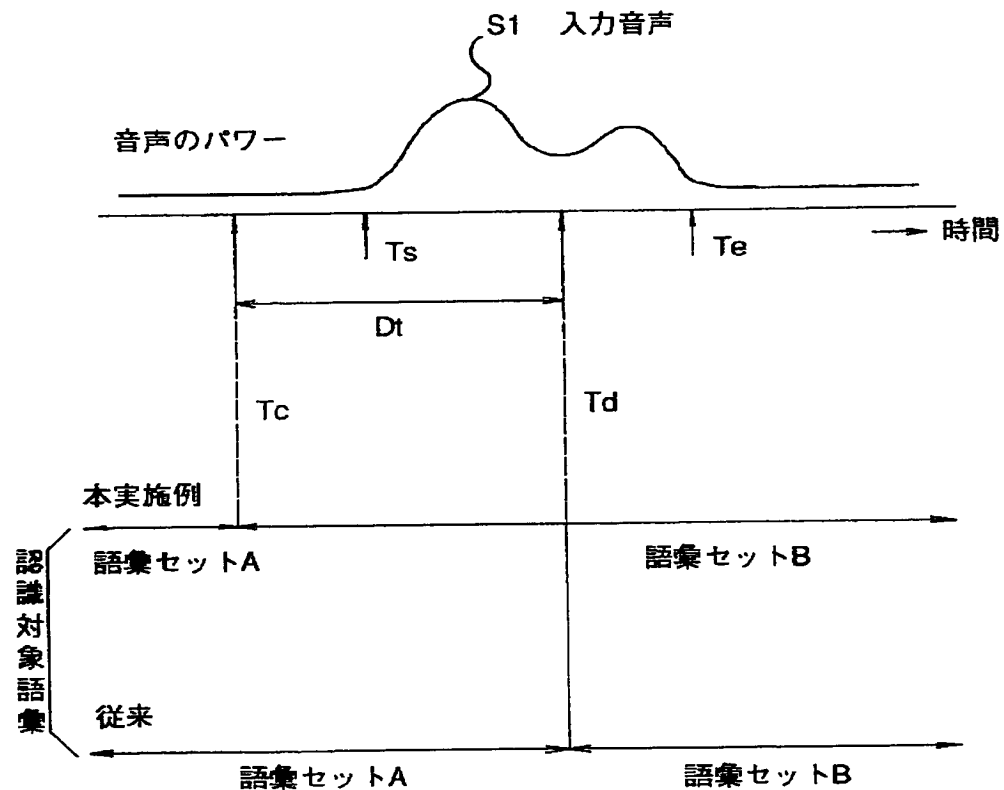
【図3】



【図2】



【図4】



**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☒ **FADED TEXT OR DRAWING**
- ☒ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☒ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

THIS PAGE BLANK (USPTO)